**Aim: WordCount Application using Java and MapReduce**

**Steps for WordCount**

Step 1 : Run your cloudera system. Open Eclipse.



Step 2 : Click on File > New > java project. Give Project Name (“WordCount”).



Step 3 : Click on Libraries tab. Then click on Add External JARs… Tap To add Hadoop Libraries.



Step 4 : Follow this steps : Click on File System -> usr -> lib -> hadoop (Select all the libraries (JAR files) -> Click OK.

Again Click on Add External JARs… -> client -> select all jar files -> ok -> Finish.



Step : 5



Step : 6



Step : 7



Step : 8



Step : 9



Step : 10



Step 11 : Now you will see project name “WordCount” on sidebar. Right Click on Project name “WordCount” -> New -> class.

For package write name “WordCloud” -> Finish. WordCount.java window will pop up.



Step : 12



Step 13 : Write your code in window(WordCount.java).



Step 14 : Right Click on the project name

WordCount -> Export -> Java -> JAR File -> Next -> For select the export destination for JAR file:

browse -> Name : WordCount.jar -> save in folder -> cloudera -> Finish -> OK



Step : 15



Step : 16



Step 17 : Now open terminal and follow this commands.

[cloudera@quickstart ~]$ ls

cloudera-manager Desktop Downloads enterprise-deployment.json kerberos Music Pictures Templates workspace

cm\_api.py Documents eclipse express-deployment.json lib parcels Public Videos

[cloudera@quickstart ~]$ cd Desktop

[cloudera@quickstart Desktop]$ mkdir wordcount

[cloudera@quickstart Desktop]$ cd wordcount

[cloudera@quickstart wordcount]$ ls

[cloudera@quickstart wordcount]$ gedit abc.txt

[cloudera@quickstart wordcount]$ cat abc.txt

Hii, my name is Siddhesh.

I am studying in RJ College

[cloudera@quickstart wordcount]$ hdfs dfs -ls /

Found 8 items

drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks

drwxr-xr-x - cloudera supergroup 0 2023-02-18 00:53 /char\_inputdir

drwxr-xr-x - cloudera supergroup 0 2023-02-18 01:01 /char\_outputdir

drwxr-xr-x - hbase supergroup 0 2023-02-25 05:53 /hbase

drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr

drwxrwxrwt - hdfs supergroup 0 2023-02-03 02:48 /tmp

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var

[cloudera@quickstart wordcount]$ hdfs dfs -mkdir /inputdir

[cloudera@quickstart wordcount]$ hdfs dfs -ls /

Found 9 items

drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks

drwxr-xr-x - cloudera supergroup 0 2023-02-18 00:53 /char\_inputdir

drwxr-xr-x - cloudera supergroup 0 2023-02-18 01:01 /char\_outputdir

drwxr-xr-x - hbase supergroup 0 2023-02-25 05:53 /hbase

drwxr-xr-x - cloudera supergroup 0 2023-02-25 06:49 /inputdir

drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr

drwxrwxrwt - hdfs supergroup 0 2023-02-03 02:48 /tmp

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var

[cloudera@quickstart wordcount]$ hdfs dfs -put /home/cloudera/Desktop/wordcount/abc.txt /inputdir/

[cloudera@quickstart wordcount]$ hdfs dfs -ls /inputdir

Found 1 items

-rw-r--r-- 1 cloudera supergroup 54 2023-02-25 06:51 /inputdir/abc.txt

[cloudera@quickstart wordcount]$ hdfs dfs -cat /inputdir/abc.txt

Hii, my name is Siddhesh.

I am studying in RJ College

[cloudera@quickstart wordcount]$ hadoop jar /home/cloudera/Desktop/WordCloud.jar WordCount /inputdir/abc.txt /outputsdir

23/02/25 06:54:31 INFO mapreduce.Job: map 0% reduce 0%

23/02/25 06:54:47 INFO mapreduce.Job: map 100% reduce 0%

23/02/25 06:55:06 INFO mapreduce.Job: map 100% reduce 100%

23/02/25 06:55:07 INFO mapreduce.Job: Job job\_1677333197123\_0001 completed successfully

23/02/25 06:55:08 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=126

FILE: Number of bytes written=286941

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=167

HDFS: Number of bytes written=76

HDFS: Number of read operations=6

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

Job Counters

Launched map tasks=1

Launched reduce tasks=1

Data-local map tasks=1

Total time spent by all maps in occupied slots (ms)=13670

Total time spent by all reduces in occupied slots (ms)=16088

Total time spent by all map tasks (ms)=13670

Total time spent by all reduce tasks (ms)=16088

Total vcore-milliseconds taken by all map tasks=13670

Total vcore-milliseconds taken by all reduce tasks=16088

Total megabyte-milliseconds taken by all map tasks=13998080

Total megabyte-milliseconds taken by all reduce tasks=16474112

Map-Reduce Framework

Map input records=2

Map output records=11

Map output bytes=98

Map output materialized bytes=126

Input split bytes=113

Combine input records=11

Combine output records=11

Reduce input groups=11

Reduce shuffle bytes=126

Reduce input records=11

Reduce output records=11

Spilled Records=22

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=451

CPU time spent (ms)=1150

Physical memory (bytes) snapshot=360525824

Virtual memory (bytes) snapshot=3015397376

Total committed heap usage (bytes)=223350784

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=54

File Output Format Counters

Bytes Written=76

[cloudera@quickstart wordcount]$ hdfs dfs -ls /

Found 10 items

drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks

drwxr-xr-x - cloudera supergroup 0 2023-02-18 00:53 /char\_inputdir

drwxr-xr-x - cloudera supergroup 0 2023-02-18 01:01 /char\_outputdir

drwxr-xr-x - hbase supergroup 0 2023-02-25 05:53 /hbase

drwxr-xr-x - cloudera supergroup 0 2023-02-25 06:51 /inputdir

drwxr-xr-x - cloudera supergroup 0 2023-02-25 06:55 /outputsdir

drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr

drwxrwxrwt - hdfs supergroup 0 2023-02-03 02:48 /tmp

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var

[cloudera@quickstart wordcount]$ hdfs dfs -ls /outputsdir

Found 2 items

-rw-r--r-- 1 cloudera supergroup 0 2023-02-25 06:55 /outputsdir/\_SUCCESS

-rw-r--r-- 1 cloudera supergroup 76 2023-02-25 06:55 /outputsdir/part-r-00000

[cloudera@quickstart wordcount]$ hdfs dfs -cat /outputsdir/part-r-00000

**Output:**

College 1

Hii, 1

I 1

RJ 1

Siddhesh. 1

am 1

in 1

is 1

my 1

name 1

studying 1